1 极大似然估计 ML

极大似然估计 (ML) 是用来估计统计模型的参数。

- x : 观测数据;
- θ : 模型参数;
- $p(x;\theta)$: 作为参数 θ 的函数，称作似然函数。

极大似然估计是说既然数据已经存在，那么模型的参数要使得数据的概率 $p(x;\theta)$ 最大:

$$\hat{\theta}_{ML} = \arg\max_{\theta} p(x;\theta) \tag{1}$$

2 期望最大化 EM

EM 算法是用来求解含有隐变量的统计模型的一种方法。

- x : 观测数据
- θ : 模型参数
- z : 隐变量

2.1 基于 Jensen 不等式的推导

$$\begin{aligned} \log p(x;\theta) &= \log\left[\sum_z p(x,z;\theta)\right] \\ &= \log\left[\sum_z q(z)\frac{p(x,z;\theta)}{q(z)}\right] \\ &\geq \sum_z \left[q(z)\log\frac{p(x,z;\theta)}{q(z)}\right] \end{aligned} \tag{2}$$

当 $q(z) = p(z|x;\theta)$ 时，等式成立。

EM 算法有 E-step 和 M-step。E-step 是给定参数 θ^t，求隐变量的分布 $q(z) = p(z|x;\theta^t)$;M-step 是将求

出来的隐变量的分布 $q(z)=p(z|x;\theta^t)$ 代入公式 (2) 的右边得到 $\log p(x;\theta)$ 的一个下界函数:

$$
\begin{aligned}
g_t(\theta) &= \sum_z \left[p(z|x;\theta^t)\log\frac{p(x,z;\theta)}{p(z|x;\theta^t)} \right] \\
&= \sum_z \left[p(z|x;\theta^t)\log p(x,z;\theta) \right] - \sum_z \left[p(z|x;\theta^t)\log p(z|x;\theta^t) \right] \\
&= \mathcal{Q}(\theta|\theta^t) + \mathcal{H}(q) \\
&= \mathcal{Q}(\theta|\theta^t) + constant
\end{aligned} \tag{3}
$$

其中

$$
\mathcal{Q}(\theta|\theta^t) = \sum_z \left[p(z|x;\theta^t)\log p(x,z;\theta) \right] \tag{4}
$$

以及 $\mathcal{H}(q)=\sum_z \left[p(z|x;\theta^t)\log p(z|x;\theta^t) \right]$。

总结如下:

E-step: 给定 θ^t，根据公式 (2) 求出隐变量的分布:

$$
q^{t+1}(z) = p(z|x;\theta^t) \tag{5}
$$

M-step: 根据 E-step 的隐变量分布 $q^{t+1}(z)=p(z|x;\theta^t)$, 最大化 $g_t(\theta)$ 来求解 θ, 公式 (3) 说明了最大化 $g_t(\theta)$ 只需要最大化公式 (4) 的 $\mathcal{Q}(\theta|\theta^t)$:

$$
\theta^{t+1} = \arg\max_\theta \mathcal{Q}(\theta|\theta^t) \tag{6}
$$

2.2 基于变分推断框架的推导

2.2.1 一个等式

从如下等式开始:

$$
p(x;\theta) = \frac{p(x,z;\theta)}{p(z|x;\theta)} \tag{7}
$$

对 (7) 式右边的分式的分子与分母同除以 $q(z)$ 后取对数:

$$
\log p(x;\theta) = \log\frac{p(x,z;\theta)}{q(z)} - \log\frac{p(z|x;\theta)}{q(z)} \tag{8}
$$

在 (8) 式两边对 $q(z)$ 求期望:

$$
E_q \log p(x;\theta) = E_q \log\frac{p(x,z;\theta)}{q(z)} - E_q \log\frac{p(z|x;\theta)}{q(z)} \tag{9}
$$

由于 $E_q \log p(x;\theta) = \log p(x;\theta)$，得下式:

$$\log p(x;\theta) = \mathcal{F}(q,\theta) + \mathcal{KL}(q||p) \tag{10}$$

其中

$$\mathcal{F}(q,\theta) = E_q \log \frac{p(x,z;\theta)}{q(z)} = \sum_z \left[q(z) \log \frac{p(x,z;\theta)}{q(z)} \right] \tag{11}$$

以及

$$\mathcal{KL}(q||p) = -E_q \log \frac{p(z|x;\theta)}{q(z)} = \sum_z \left[q(z) \log \frac{q(z)}{p(z|x;\theta)} \right] \tag{12}$$

2.2.2 E-step 和 M-step

E-step: 给定 θ^t，我们寻找 $q(z)$ 使得**证据下界** $\mathcal{F}(q,\theta)$ 最大:

$$\begin{aligned} q^{t+1}(z) &= \arg\max_q \mathcal{F}(q,\theta^t) \\ &= \arg\min_q \mathcal{KL}(q||p(z|x;\theta^t) \\ &= p(z|x;\theta^t) \end{aligned} \tag{13}$$

Remark:$\mathcal{F}(q,\theta)$ 又被称为**变分自由能**。

M-step: 我们寻找模型参数 θ 使得 $F(q^{t+1}(z),\theta)$ 最大:

$$\begin{aligned} \theta^{t+1} &= \arg\max_\theta \mathcal{F}(q^{t+1}(z),\theta) \\ &= \arg\max_\theta E_{q^{t+1}(z)} \log p(x,z;\theta) \end{aligned} \tag{14}$$

Remark:$\mathcal{Q}(\theta|\theta^t) = E_{q^{t+1}(z)} \log p(x,z;\theta)$